Sitemaps

This is the latest accepted revision, reviewed on *9 April 2025*.

Sitemaps is a protocol in XML format meant for a webmaster to inform search engines about URLs on a website that are available for web crawling. It allows webmasters to include additional information about each URL: when it was last updated, how often it changes, and how important it is in relation to other URLs of the site. This allows search engines to crawl the site more efficiently and to find URLs that may be isolated from the rest of the site's content. The Sitemaps protocol is a URL inclusion protocol and complements `robots.txt`, a URL exclusion protocol.

History

Google first introduced Sitemaps 0.84 in June 2005 so web developers could publish lists of links from across their sites.[1] Google, Yahoo! and Microsoft announced joint support for the Sitemaps protocol in November 2006.[2] The schema version was changed to "Sitemap 0.90", but no other changes were made.

In April 2007, Ask.com and IBM announced support for Sitemaps.[3] Also, Google, Yahoo, MSN announced auto-discovery for sitemaps through `robots.txt`. In May 2007, the state governments of Arizona, California, Utah and Virginia announced they would use Sitemaps on their web sites.[4]

The Sitemaps protocol is based on ideas[5] from "Crawler-friendly Web Servers,"[6] with improvements including auto-discovery through `robots.txt` and the ability to specify the priority and change frequency of pages.

Purpose

Sitemaps are particularly beneficial on websites where:

- Some areas of the website are not available through the browsable interface[7]
- Webmasters use rich Ajax, Silverlight, or Flash content that is not normally processed by search engines.
- The site is very large and there is a chance for the web crawlers to overlook some of the new or recently updated content[7]
- When websites have a huge number of pages that are isolated or not well linked together, or[7]
- When a website has few external links[7]

File format

The Sitemap Protocol format consists of XML tags. The file itself must be UTF-8 encoded. Sitemaps can also be just a plain text list of URLs. They can also be compressed in .gz format.

A sample Sitemap that contains just one URL and uses all optional tags is shown below.

```
<?xml version='1.0' encoding='UTF-8'?>
<urlset xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd"
    xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
    <url>
        <loc>https://example.com/</loc>
        <lastmod>2006-11-18</lastmod>
        <changefreq>daily</changefreq>
        <priority>0.8</priority>
    </url>
</urlset>
```

The Sitemap XML protocol is also extended to provide a way of listing multiple Sitemaps in a 'Sitemap index' file. The maximum Sitemap size of 50 MiB or 50,000 URLs[8] means this is necessary for large sites.

An example of Sitemap index referencing one separate sitemap follows.

```
<?xml version="1.0" encoding="UTF-8"?>
<sitemapindex xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
http://www.sitemaps.org/schemas/sitemap/0.9/siteindex.xsd"
    xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
   <sitemap>
      <loc>https://www.example.com/sitemap1.xml.gz</loc>
      <lastmod>2014-10-01T18:23:17+00:00</lastmod>
   </sitemap>
</sitemapindex>
```

Element definitions

The definitions for the elements are shown below:[8]

Element	Required?	Description
`<urlset>`	Yes	The document-level element for the Sitemap. The rest of the document after the '<?xml version>' element must be contained in this.
`<url>`	Yes	Parent element for each entry.
`<sitemapindex>`	Yes	The document-level element for the Sitemap index. The rest of the document after the '<?xml version>' element must be contained in this.
`<sitemap>`	Yes	Parent element for each entry in the index.
`<loc>`	Yes	Provides the full URL of the page or sitemap, including the protocol (e.g. http, https) and a trailing slash, if required by the site's hosting server. This value must be shorter than 2,048 characters. Note that ampersands in the URL need to be escaped as `&`.
`<lastmod>`	No	The date that the file was last modified, in ISO 8601 format. This can display the full date and time or, if desired, may simply be the date in the format YYYY-MM-DD.
`<changefreq>`	No	How frequently the page may change: • always • hourly • daily • weekly • monthly • yearly • never "Always" is used to denote documents that change each time that they are accessed. "Never" is used to denote archived URLs (i.e. files that will not be changed again). This is used only as a guide for crawlers, and is not used to determine how frequently pages are indexed. Does not apply to `<sitemap>` elements.
`<priority>`	No	The priority of that URL relative to other URLs on the site. This allows webmasters to suggest to crawlers which pages are considered more important. The valid range is from 0.0 to 1.0, with 1.0 being the most important. The default value is 0.5. Rating all pages on a site with a high priority does not affect search listings, as it is only used to suggest to the crawlers how important pages of the site are to one another.

		Does not apply to `<sitemap>` elements.

Support for the elements that are not required can vary from one search engine to another.[8]

Other formats

Text file

The Sitemaps protocol allows the Sitemap to be a simple list of URLs in a text file. The file specifications of XML Sitemaps apply to text Sitemaps as well; the file must be UTF-8 encoded, and cannot be more than 50MiB (uncompressed) or contain more than 50,000 URLs. Sitemaps that exceed these limits should be broken up into multiple sitemaps with a sitemap index file (a file that points to multiple sitemaps).[9]

Syndication feed

A syndication feed is a permitted method of submitting URLs to crawlers; this is advised mainly for sites that already have syndication feeds. One stated drawback is this method might only provide crawlers with more recently created URLs, but other URLs can still be discovered during normal crawling.[8]

It can be beneficial to have a syndication feed as a delta update (containing only the newest content) to supplement a complete sitemap.

Search engine submission

If Sitemaps are submitted directly to a search engine (pinged), it will return status information and any processing errors. The details involved with submission will vary with the different search engines. The location of the sitemap can also be included in the `robots.txt` file by adding the following line:

```
Sitemap: <sitemap_location>
```

The `<sitemap_location>` should be the complete URL to the sitemap, such as:

```
https://www.example.org/sitemap.xml
```

This directive is independent of the user-agent line, so it doesn't matter where it is placed in the file. If the website has several sitemaps, multiple "Sitemap:" records may be included in `robots.txt`, or the URL can simply point to the main sitemap index file.

The following table lists the sitemap submission URLs for a few major search engines:

Search engine	Submission URL	Help page	Market
Baidu	**https://zhanzhang.baidu.com/dashboard/index**	Baidu Webmaster Dashboard (https://zhanzhang.baidu.com/sitemap/index)	China, Singapore
Bing (and Yahoo!)	**https://www.bing.com/webmaster/ping.aspx?siteMap=**	Bing Webmaster Tools (https://www.bing.com/webmaster)	Global
Yandex	**https://webmaster.yandex.com/site/map.xml**	Sitemaps files (https://help.yandex.com/webmaster/?id=1115259)	Russia, Belarus, Kazakhstan, Turkey

Sitemap URLs submitted using the sitemap submission URLs need to be URL-encoded, for example: replace `:` (colon) with `%3A`, replace `/` (slash) with `%2F`.[8]

Google retired sitemap submissions using URLs in late 2023.[10]

Limitations for search engine indexing

Sitemaps supplement and do not replace the existing crawl-based mechanisms that search engines already use to discover URLs. Using this protocol does not guarantee that web pages will be included in search indexes, nor does it influence the way that pages are ranked in search results. Specific examples are provided below.

- Google - Webmaster Support on Sitemaps: "Using a sitemap doesn't guarantee that all the items in your sitemap will be crawled and indexed, as Google processes rely on complex algorithms to schedule crawling. However, in most cases, your site will benefit from having a sitemap, and you'll never be penalized for having one."[11]
- Bing - Bing uses the standard sitemaps.org protocol and is very similar to the one mentioned below.
- Yahoo - After the search deal commenced between Yahoo! Inc. and Microsoft, Yahoo! Site Explorer has merged with Bing Webmaster Tools (https://web.archive.org/web/20110417233919/http://siteexplorer.search.yahoo.com/)

Sitemap limits

Sitemap files have a limit of 50,000 URLs and 50MiB (52,428,800 bytes) per sitemap. Sitemaps can be compressed using gzip, reducing bandwidth consumption. Multiple sitemap files are supported, with a Sitemap index file serving as an entry point. Sitemap index files may not list

more than 50,000 Sitemaps and must be no larger than 50MiB and can be compressed. You can have more than one Sitemap index file.[8]

As with all XML files, any data values (including URLs) must use entity escape codes for the characters ampersand (&), single quote ('), double quote ("), less than (<), and greater than (>).

Best practice for optimising a sitemap index for search engine crawlability is to ensure the index refers only to sitemaps as opposed to other sitemap indexes. Nesting a sitemap index within a sitemap index is invalid according to Google.[12]

Additional sitemap types

A number of additional XML sitemap types outside of the scope of the Sitemaps protocol are supported by Google to allow webmasters to provide additional data on the content of their websites. Video and image sitemaps are intended to improve the capability of websites to rank in image and video searches.[13][14]

Video sitemaps

Video sitemaps indicate data related to embedding and autoplaying, preferred thumbnails to show in search results, publication date, video duration, and other metadata.[14] Video sitemaps are also used to allow search engines to index videos that are embedded on a website, but that are hosted externally, such as on Vimeo or YouTube.

Image sitemaps

Image sitemaps are used to indicate image metadata, such as licensing information, geographic location, and an image's caption.[13]

Google News Sitemaps

Google supports a Google News sitemap type for facilitating quick indexing of time-sensitive news subjects.[15][16]

Multilingual and multinational sitemaps

In December 2011, Google announced the annotations for sites that want to target users in many languages and, optionally, countries. A few months later Google announced, on their official blog,[17] that they are adding support for specifying the rel="alternate" and hreflang annotations in

Sitemaps. Instead of the (until then only option) HTML link elements the Sitemaps option offered many advantages which included a smaller page size and easier deployment for some websites.

One example of the multilingual sitemap would be as follows:

If for example we have a site that targets English language users through `https://www.example.com/en` and Greek language users through `https://www.example.com/gr`, up until then the only option was to add the hreflang annotation either in the HTTP header or as HTML elements on both URLs like this

```
<link rel="alternate" hreflang="en"
href="https://www.example.com/en" />
<link rel="alternate" hreflang="gr"
href="https://www.example.com/gr" />
```

But now, one can alternatively use the following equivalent markup in Sitemaps:

```
<url>
  <loc>https://www.example.com/en</loc>
   <xhtml:link
     rel="alternate"
     hreflang="gr"
     href="https://www.example.com/gr" />
   <xhtml:link
     rel="alternate"
     hreflang="en"
     href="https://www.example.com/en" />
</url>
<url>
  <loc>https://www.example.com/gr</loc>
   <xhtml:link
     rel="alternate"
     hreflang="gr"
     href="https://www.example.com/gr" />
   <xhtml:link
     rel="alternate"
     hreflang="en"
     href="https://www.example.com/en" />
</url>
```

See also

- Biositemap
- Metadata

- Resources of a Resource
- Yahoo! Site Explorer
- Google Webmaster Tools

References

1. Shivakumar, Shiva (2005-06-02). "Google Blog: Webmaster-friendly" (http://googleblog.blogspot.com/2005/06/webmaster-friendly.html) . Archived (https://web.archive.org/web/20050608015054/http://googleblog.blogspot.com/2005/06/webmaster-friendly.html) from the original on 2005-06-08. Retrieved 2021-12-31.
2. "Major Search Engines Unite to Support a Common Mechanism for Website Submission" (https://googlepress.blogspot.com/2006/11/major-search-engines-unite-to-support_16.html) . *News from Google*. November 16, 2006. Retrieved 2021-12-31.
3. Pathak, Vivek (2007-05-11). "The Ask.com Blog: Sitemaps Autodiscovery" (https://web.archive.org/web/20070518162623/http://blog.ask.com/2007/04/sitemaps_autodi.html) . *Ask's Official Blog*. Archived from the original (http://blog.ask.com/2007/04/sitemaps_autodi.html) on 2007-05-18. Retrieved 2021-12-31.
4. "Information for Public Sector Organizations" (https://web.archive.org/web/20070430123308/http://www.google.com/publicsector/) . Archived from the original (http://www.google.com/publicsector/) on 2007-04-30.
5. M.L. Nelson; J.A. Smith; del Campo; H. Van de Sompel; X. Liu (2006). "Efficient, Automated Web Resource Harvesting" (http://www.cs.odu.edu/~mln/pubs/widm-2006/modoai-widm06.pdf) (PDF). *WIDM'06*.
6. O. Brandman, J. Cho, Hector Garcia-Molina, and Narayanan Shivakumar (2000). "Crawler-friendly web servers". *Proceedings of ACM SIGMETRICS Performance Evaluation Review, Volume 28, Issue 2*. doi:10.1145/362883.362894 (https://doi.org/10.1145%2F362883.362894) .
7. "Learn about sitemaps | Search Central" (https://developers.google.com/search/docs/advanced/sitemaps/overview) . *Google Developers*. Retrieved 2021-06-01.
8. "Sitemaps XML format" (https://www.sitemaps.org/protocol.html) . Sitemaps.org. 2016-11-21. Retrieved 2016-12-01.
9. "Build and submit a sitemap - Search Console Help" (https://support.google.com/webmasters/bin/answer.py?hl=en&answer=183668) . *Support.google.com*. Retrieved 30 November 2020.

10. "Sitemaps ping endpoint is going away" (https://developers.google.com/search/blog/2023/06/sitemaps-lastmod-ping) . 2025-04-04. Retrieved 2025-04-04.
11. "About Google Sitemaps" (https://support.google.com/webmasters/answer/156184?hl=en) . 2016-12-01. Retrieved 2016-12-01.
12. "Sitemaps report - Search Console Help" (https://support.google.com/webmasters/answer/7451001?visit_id=636854840848822341-3070733684&rd=2) . *support.google.com*. Retrieved 2020-04-15.
13. "Image Sitemaps" (https://support.google.com/webmasters/answer/80471?hl=en&ref_topic=4581190) . *Google Search Console*. Retrieved 28 December 2018.
14. "Video Sitemaps" (https://support.google.com/webmasters/answer/80471?hl=en&ref_topic=4581190) . *Google Search Console*. Retrieved 28 December 2018.
15. Bigby, Garenne. "Why You should be using a Google News Sitemap" (https://dynomapper.com/blog/22-sitemaps/205-why-you-should-be-using-a-google-news-sitemap) . *Dyno Mapper*. Retrieved 28 December 2018.
16. "Google News Sitemaps" (https://support.google.com/webmasters/answer/80471?hl=en&ref_topic=4581190) . *Google Search Console*. Retrieved 28 December 2018.
17. "Multilingual and multinational site annotations in Sitemaps" (http://googlewebmastercentral.blogspot.gr/2012/05/multilingual-and-multinational-site.html) . *Google Webmaster Central Blog*. Pierre Far. May 24, 2012.

External links

- Official website (https://www.sitemaps.org/)
- Google news groups
  - Sitemaps (https://developers.google.com/search/docs/crawling-indexing/sitemaps/overview)
  - Webmaster help - Sitemap (https://groups.google.com/group/Google_Webmaster_Help-Sitemap) Archived (https://web.archive.org/web/20061221154105/http://groups.google.com/group/Google_Webmaster_Help-Sitemap) 2006-12-21 at the Wayback Machine